for Comparisons of Two-Dimensional Maps of Proteins from Different Human Cell Types Defined in a pH Scale Where Isoelectric Points Correlate with Polypeptide Compositions, Electrophoresis, 15, 529-539 (1994) (hereinafter "the Bjellqvist article") (copy annexed at Tab F);

(g) Large Scale Biology Company Info; LSB and LSP Information; from http://www.lsbc.com (2001) (copy annexed at Tab G);

(h) Celio, M.R., Pauls, T., and Schwaller, B., Guidebook to the Calcium-Binding Proteins, Oxford University Press, Oxford, UK, pp. 15-20 (1996) (hereinafter "the Celio article") (copy annexed at Tab H).

(i) Schwaninger, M., Blume, R., Oetjen, E., Lux, G., and Knepel, W., Inhibition of cAMP-responsive Element-mediated Gene Transcription by Cyclosporin A and FK506 after Membrane Depolarization, Journal of Biological Chemistry, 268, 23111-23115 (1993) (hereinafter "the Schwaninger article") (copy annexed at Tab I).

(j) Rasmussen, C. D., and Means, A.R., Calmodulin, cell growth and gene expression, Trends in Neurosciences, 12, 432-438 (1989) (hereinafter, "the Rasmussen article") (copy annexed in Tab J).

8. Many of the published articles I considered (i.e., at least items (a)-(f) identified in paragraph 7) relate to the development of protein two-dimensional gel electrophoretic techniques for use in gene expression monitoring applications in drug development and toxicology. As I will discuss below, a person skilled in the art who read the Hillman '409 application on November 3, 1997 would have understood that application to disclose the SEQ ID NO:1 polypeptide to be useful for a number of gene and protein expression monitoring applications, e.g., in the use of two-dimensional polyacrylamide gel electrophoresis and western blot analysis of tissue samples in drug development and in toxicity testing.

9. Turning more specifically to the Hillman '409 specification, the SEQ ID NO:1 polypeptide is shown at pages 53-54 as one of 5 sequences under the heading "Sequence Listing." The Hillman '409 specification specifically teaches that the "invention features a substantially purified polypeptide, calmodulin protein (DACP-1), having the amino acid sequence shown in SEQ ID NO:1" (Hillman '409 application at p. 2). It further teaches that (a) the identity of the SEQ ID NO:1 polypeptide was determined from a "breast tumor cDNA library" (BRSTTUT14), (b) the SEQ ID NO:1 polypeptide is the calmodulin protein referred to

as "DACP-1" and is encoded by SEQ ID NO:2, and (c) northern analysis (Figure 3) shows that DACP-1 "is expressed in various libraries, at least 73% of which are immortalized or cancerous, at least 55% of which involve the reproductive system, and at least 27% of which involve the immune response" and therefore "DACP-1 appears to play a role in cancer and immune and reproductive disorders" (Hillman '409 application at pp. 13-14, 25, and 43).

The Hillman '409 application discusses a number of uses of the SEQ ID NO:1 polypeptide in addition to its use in gene and protein expression monitoring applications. I have not fully evaluated these additional uses in connection with the preparation of this Declaration and do not express any views in this Declaration regarding whether or not the Hillman '409 specification discloses these additional uses to be substantial, specific and credible real-world utilities of the SEQ ID NO:1 polypeptide. Consequently, my discussion in this Declaration concerning the Hillman '409 application focuses on the portions of the application that relate to the use of the SEQ ID NO:1 polypeptide in gene and protein expression monitoring applications.

10. The Hillman '409 application discloses that the polynucleotide sequences disclosed therein, including the polynucleotides encoding the SEQ ID NO:1 polypeptide, are useful as probes in chip based technologies. It further teaches that the chip based technologies can be used "for the detection and/or quantification of nucleic acid or protein" (Hillman '409 application at p. 23).

The Hillman '409 application also discloses that the SEQ ID NO:1 polypeptide is useful in other protein expression detection technologies. The Hillman '409 application states that "[a] variety of protocols for detecting and measuring the expression of DACP-1, using either polyclonal or monoclonal antibodies specific for the protein are known in the art. Examples include enzyme-linked immunosorbent assay (ELISA), radioimmunoassay (RIA), and fluorescence activated cell sorting (FACS)" (Hillman '409 application at p. 23, ¶ 4). Furthermore, the Hillman '409 application discloses that "[a] variety of protocols including ELISA, RIA, and FACS for measuring DACP-1 are known in the art and provide a basis for diagnosing altered or abnormal levels of DACP-1 expression. Normal or standard values for DACP-1 expression are established by combining body fluids or cell extracts taken from normal mammalian subjects, preferably human, with antibody to DACP-1 under conditions suitable for complex formation" (Hillman '409 application at p. 35, ¶ 2).

In addition, at the time of filing the Hillman '409 application, it was well known in the art that "gene" and protein expression analyses also included two-dimensional polyacrylamide gel electrophoresis (2-D PAGE) technologies, which were developed during the

1980s, and as exemplified by the Anderson 1991 and 1995 articles (Tab C and Tab D). The Anderson 1991 article teaches that a 2-D PAGE map has been used to connect and compare hundreds of 2-D gels of rat liver samples from a variety of studies including regulation of protein expression by various drugs and toxic agents (Tab C at p. 907). The Anderson 1991 article teaches an empirically-determined standard curve fitted to a series of identified proteins based upon amino acid chain length (Tab C at p. 911) and how that standard curve can be used in protein expression analysis. The Anderson 1991 article teaches that "there is a long-term need for a comprehensive database of liver proteins" (Tab C at p. 912).

The Wilkins article is one of a number of documents that were published prior to the November 3, 1997 filing date of the Hillman '409 application that describes the use of the 2-D PAGE technology in a wide range of gene and protein expression monitoring applications, including monitoring and analyzing protein expression patterns in human cancer, human serum plasma proteins, and in rodent liver following exposure to toxins. In view of the Hillman '409 application, the Wilkins article, and other related pre-November 1997 publications, persons skilled in the art on November 3, 1997 clearly would have understood the Hillman '409 application to disclose the SEQ ID NO:1 polypeptide to be useful in 2-D PAGE analyses for the development of new drugs and monitoring the activities of drugs or th treatment of cancer and immune disorders for such purposes as evaluating their efficacy and toxicity, as explained more fully in paragraph 13 below. With specific reference to toxicity evaluations, those of skill in the art who were working on drug development in November 1997 (and for many years prior to November 1997) without any doubt appreciated that the toxicity (or lack of toxicity) of any proposed drug they were working on was one of the most important criteria to be considered and evaluated in connection with the development of the drug. They would have understood at that time that good drugs are not only potent, they are specific. This means that they have strong effects on a specific biological target and minimal effects on all other biological targets. Ascertaining that a candidate drug affects its intended target, and identification of undesirable secondary effects (i.e., toxic side effects), had been for many years among the main challenges in developing new drugs. The ability to determine which genes are positively affected by a given drug, coupled with the ability to quickly and at the earliest time possible in the drug development process identify drugs that are likely to be toxic because of their undesirable secondary effects, have enormous value in improving the efficiency of the drug discovery process, and are an important and essential part of the development of any new drug. In fact, the desire to identify and understand toxicological effects using the experimental assays described above led Dr Leigh Anderson to found the Large Scale Biology Corporation in 1985, in order to pursue commercial

development of the 2-D electrophoretic protein mapping technology he had developed. In addition, the company focused on toxicological effects on the proteome as clearly demonstrated by its goals and by its senior management credentials described in company documents (see Tab G at pp. 1, 3, and 5).

Accordingly, the teachings in the Hillman '409 application, in particular regarding use of SEQ ID NO:1 in differential gene and protein expression analysis (2-D PAGE maps) and in the development and the monitoring of the activities of drugs, clearly includes toxicity studies and persons skilled in the art who read the Hillman '409 application on November 3, 1997 would have understood that to be so.

11. As previously discussed (*supra*, paragraphs 7 and 8), my experience with protein analysis methods in the mid-1980s and the several publications annexed to this Declaration at Tabs A through F evidence information that was available to the public regarding two-dimensional polyacrylamide gel electrophoresis technology and its uses in drug discovery and toxicology testing before the November 3, 1997 filing date of the Hillman '409 application. In particular the Celis article stated that "protein databases are expected to foster a variety of biological information.... -- among others, ... drug development and testing" (See Tab D, p. 2200, second column). The Franzen article shows that 2-D PAGE maps were used to identify proteins in clinical tumor material (See Tab E). The Hillman '409 application clearly discloses that expression of DACP-1 is associated with cancer, in particular, cancers of the reproductive system and bladder (Hillman '409 application at p. 14, lines 1-3, and Figure 3). The Bjellqvist article showed that a protein may be identified accurately by its positional co-ordinates, namely molecular mass and isoelectric point (See Tab F). The Hillman '409 application clearly disclosed SEQ ID NO:1 from which it would have been routine for one of skill in the art to predict both the molecular mass and the isoelectric point using algorithms well known in the art at the time of filing.

check '409

12. A person skilled in the art on November 3, 1997, who read the Hillman '409 application, would understand that application to disclose the SEQ ID NO:1 polypeptide to be highly useful in analysis of differential expression of proteins. For example, the specification of the Hillman '409 application would have led a person skilled in the art in November 1997 who was using protein expression monitoring in connection with working on developing new drugs for the treatment of cancer or immune disorders to conclude that a 2-D PAGE map that used the substantially purified SEQ ID NO:1 polypeptide would be a highly useful tool and to

request specifically that any 2-D PAGE map that was being used for such purposes utilize the SEQ ID NO:1 polypeptide sequence. Expressed proteins are useful for 2-D PAGE analysis in toxicology expression studies for a variety of reasons, particularly for purposes relating to providing controls for the 2-D PAGE analysis, and for identifying sequence or post-translational variants of the expressed sequences in response to exogenous compounds. Persons skilled in the art would appreciate that a 2-D PAGE map that utilized the SEQ ID NO:1 polypeptide sequence would be a more useful tool than a 2-D PAGE map that did not utilize this protein sequence in connection with conducting protein expression monitoring studies on proposed (or actual) drugs for treating cancer or immune disorders for such purposes as evaluating their efficacy and toxicity.

I discuss in more detail in items (a)-(c) below a number of reasons why a person skilled in the art, who read the Hillman '409 specification in November 1997, would have concluded based on that specification and the state of the art at that time, that SEQ ID NO:1 polypeptide would be a highly useful tool for analysis of a 2-D PAGE map for evaluating the efficacy and toxicity of proposed drugs for cancer or immune disorders by means of 2-D PAGE maps, as well as for other evaluations:

(a) The Hillman '409 specification contains a number of teachings that would lead persons skilled in the art on November 3, 1997 to conclude that a 2-D PAGE map that utilized the substantially purified SEQ ID NO:1 polypeptide would be a more useful tool for gene expression monitoring applications relating to drugs for treating cancer or immune disorders than a 2-D PAGE map that did not use the SEQ ID NO:1 polypeptide sequence. Among other things, the Hillman '409 specification teaches that (i) the identity of the SEQ ID NO:1 polypeptide was determined from a bladder tumor cDNA library, (ii) the SEQ ID NO:1 polypeptide is the calmodulin protein referred to as DACP-1, and (iii) DACP-1 is expressed in various libraries derived from cancers of the bladder, breast, prostate, and ovaries and, therefore, "DACP-1 appears to be involved in cancer and immune and reproductive disorders, (Hillman '409 application at pp. 14, 25, and Figure 3; see paragraph 9, *supra*). The substanially purified polypeptide could therefore be used as a control to more acurately guage the expression of DACP-1 in the sample and consequently more accurately gauge the affect of a toxicant on expression of the gene.

Moreover, the Hillman '409 specification teaches that SEQ ID NO:1 shares chemical and structural homology with calmodulin proteins from P. falciparum, rat and human. In particular, DACP-1 shares 51% identity with each of the three calmodulin proteins. DACP-1 further contains four potential EF-hand domains known to be characteristic of calcium-binding

proteins, and calmodulin in particular (Hillman '409 application at p. 1, line 14-19, and p. 13, lines 20-22). DACP-1 and the three calmodulin proteins share the four EF-hand domains, a potential glycosylation site, and two potential casein kinase II phosphorylation sites (Hillman '409 application, at p. 13, lines 25-30).

The specification discloses that calmodulin is the most widely distributed and common mediator of calcium-mediated cell signaling, and is involved in a multitude of cellular processes including gene regulation, DNA synthesis, cell cycle progression, etc., and that regulation of calcium-binding proteins, including calmodulins, has implications for control of a variety of disorders. In particular, calcineurin, a calmodulin-regulated protein phosphatase, is a known target for inhibition by the immunosuppresive agents cyclosporin and FK506; and calmodulin levels are increased several-fold in tumors and tumor cell lines (Hillman '409 application, p. 1, lines 20-27, and at p. 2, lines 8-14).

(b) Also pertinent is that pre-November 1997 articles cited in the Hillman '409 specification or known in the art point to the potential role in cancer and immune disorders of previously known calmodulin proteins in the same class as DACP-1. The Celio article discloses that calcium-binding proteins are characterized by the presence of four calcium-binding domains known as the "EF-hand" domain, and that EF-hand, calcium-binding proteins mediate calcium-dependent, signal transduction processes. The Celio article further discloses that the expression of several EF-hand, calcium-binding proteins is modified in various human diseases, in particular, that the level of calmodulin is increased several-fold in tumor cells and tumor tissues (Tab H, at pp. 16, 18, and 19). The Schwaninger article discloses that the powerful immunosuppressive drugs, cyclosporin and FK506, act by inhibiting the calcium/calmodulin-dependent phosphatase calcineurin, thereby implicating calmodulin ativity in these processes as well (Tab I, at p. 2311). The Rasmussen article discloses that calmodulin is implicated in the control of cell proliferation and that cells transformed with a variety of oncogenic agents have elevated levels of calmodulin (Tab J at p. 433).

(c) Persons skilled in the art on November 3, 1997 would have appreciated (i) that the protein expression monitoring results obtained using a 2-D PAGE map that utilized a SEQ ID NO:1 polypeptide would vary, depending on the particular drug being evaluated, and (ii) that such varying results would occur both with respect to the results obtained from the SEQ ID NO:1 polypeptide and from the 2-D PAGE map as a whole (including all its other individual proteins). These kinds of varying results, depending on the identity of the drug being tested, in no way detracts from my conclusion that persons skilled in the art on November 3, 1997, having read the Hillman '409 specification, would specifically request that any 2-D PAGE map that was

being used for conducting protein expression monitoring studies on drugs for treating cancer or immune disorders (*e.g.*, a toxicology study or any efficacy study of the type that typically takes place in connection with the development of a drug) utilize the SEQ ID NO:1 polypeptide sequence. Persons skilled in the art on November 3, 1997 would have wanted their 2-D PAGE map to utilize the SEQ ID NO:1 polypeptide sequence because a 2-D PAGE map that utilized protein sequence information the polypeptide (as compared to one that did not) would provide more useful results in the kind of gene and protein expression monitoring studies using 2-D PAGE maps that persons skilled in the art have been doing since well prior to November 3, 1997.

The foregoing is not intended to be an all-inclusive explanation of all my reasons for reaching the conclusions stated in this paragraph 13, and in paragraph 6, *supra*. In my view, however, it provides more than sufficient reasons to justify my conclusions stated in paragraph 6 of this Declaration regarding the Hillman '409 application disclosing to persons skilled in the art at the time of its filing substantial, specific and credible real-world utilities for the SEQ ID NO:1 polypeptide.

13. Also pertinent to my considerations underlying this Declaration is the fact that the Hillman '409 disclosure regarding the uses of the SEQ ID NO:1 polypeptide for protein expression monitoring applications is <u>not</u> limited to the use of that protein in 2-D PAGE maps. For one thing, the Hillman '409 disclosure regarding the technique used in gene and protein expression monitoring applications is broad (Hillman '409 application at, e.g., p. 23, lines 9-14 and at p. 35, lines 8-13).

In addition, the Hillman '409 specification repeatedly teaches that the protein described therein (including the SEQ ID NO:1 polypeptide) may desirably be used in any of a number of long established "standard" techniques, such as ELISA or western blot analysis, for conducting protein expression monitoring studies. See, e.g.:

(a) Hillman '409 application at p. 23, lines 20-23 ("A variety of protocols for detecting and measuring the expression of DACP-1, using either polyclonal or monoclonal antibodies specific for the protein are known in the art. Examples include enzyme-linked immunosorbent assay (ELISA), radioimmunoassay (RIA), and fluorescence activated cell sorting (FACS)");

(b) Hillman '409 application at p. 35, lines 14-22 ("A variety of protocols including ELISA, RIA, and FACS for measuring DACP-1 are known in the art and provide a basis for diagnosing altered or abnormal levels of DACP-1 expression. Normal or standard

values for DACP-1 expression are established by combining body fluids or cell extracts taken from normal mammalian subjects, preferably human, with antibody to DACP-1 under conditions suitable for complex formation. The amount of standard complex formation may be quantified by various methods, but preferably by photometric, means. Quantities of DACP-1 expressed in subject, control and disease, samples from biopsied tissues are compared with the standard values. Deviation between standard and subject values establishes the parameters for diagnosing disease").

Thus a person skilled in the art on November 3, 1997, who read the Hillman '409 specification, would have routinely and readily appreciated that the SEQ ID NO:1 polypeptide disclosed therein would be useful to conduct gene expression monitoring analyses using 2-D PAGE mapping or western blot analysis or any of the other traditional membrane-based protein expression monitoring techniques that were known and in common use many years prior to the filing of the Hillman '409 application. For example, a person skilled in the art in November 1997 would have routinely and readily appreciated that the SEQ ID NO:1 polypeptide would be a useful tool in conducting protein expression analyses, using the 2-D PAGE mapping or western analysis techniques, in furtherance of (a) the development of drugs for the treatment of cancer and immune disorders and (b) analyses of the efficacy and toxicity of such drugs.

14. I declare further that all statements made herein of my own knowledge are true and that all statements made herein on information and belief are believed to be true; and further, that these statements were made with the knowledge that willful false statements and the like so made are punishable by fine or imprisonment, or both, and that willful false statements may jeopardize the validity of this application and any patent issuing thereon.

L. Michael Furness, B.Sc.

Signed at Cambridge, United Kingdom
this 17th day of November, 2001

N. Leigh Anderson
Ricardo Esquer-Blasco
Jean-Paul Hofmann
Norman G. Anderson

Large Scale Biology Corporation, Rockville, MD

A two-dimensional gel database of rat liver proteins useful in gene regulation and drug effects studies

A standard two-dimensional (2-D) protein map of Fischer 344 rat liver (F344MST3) is presented, with a tabular listing of more than 1200 protein species. Sodium dodecyl sulfate (SDS) molecular mass and isoelectric point have been established, based on positions of numerous internal standards. This map has been used to connect and compare hundreds of 2-D gels of rat liver samples from a variety of studies, and forms the nucleus of an expanding database describing rat liver proteins and their regulation by various drugs and toxic agents. An example of such a study, involving regulation of cholesterol synthesis by cholesterol-lowering drugs and a high-cholesterol diet, is presented. Since the map has been obtained with a widely used and highly reproducible 2-D gel system (the Iso-Dalt® system), it can be directly related to an expanding body of work in other laboratories.

Contents

1 Introduction

High-resolution two-dimensional electrophoresis of proteins, introduced in 1975 by O'Farrell and others [1–4], has been used over the ensuing 16 years to examine a wide variety of biological systems, the results appearing in more than 5000 published papers. With the advent of computerized systems for analyzing two-dimensional (2-D) gel images and constructing spot databases, it is also possible to plan and assemble integrated bodies of information describing the appearance and regulation of thousands of protein gene products [5, 6]. Creating such databases involves amassing and organizing quantitative data from thousands of 2-D gels, and requires a substantial commitment in technology and resources.

Given the long-term effort required to develop a protein database, the choice of a biological system takes on considerable importance. While *in vitro* systems are ideal for answering many experimental questions, especially in cancer research and genetics, our experience with cell cultures and tissue samples suggests that some *in vivo* approaches could have major advantages. In particular, we have noticed that liver tissue samples from rats and mice appear to show greater quantitative reproducibility (in terms of individual protein expression) than replicate cell cultures. This is perhaps a natural result of the homeostasis maintained in a complete animal *vs.* the well-known variability of cell cultures, the latter due principally to differences in reagents (*e.g.*, fetal bovine serum), conditions (*e.g.*, pH) and genetic "evolution" of cell lines while in culture. It is also more difficult to generate adequate amounts of protein from cell culture systems (particularly with attached cells), forcing the investigator to resort to radioisotope-based or silver-based stain-detection methods. While these methods are more sensitive (sometimes much more sensitive) than the Coomassie Brilliant Blue (CBB) stain typically used for protein detection in "large" protein samples, they are generally more variable, more labor-intensive and, in the case of radiographic methods, may generate highly "noisy" images, due to the properties of the films used. By contrast, large protein samples can easily be prepared from liver using urea/Nonidet P-40 (NP-40) solubilization and stained with CBB, which has the advantage of being easily reproducible [8]. Finally, there remains the question of the "truthfulness" of many *in vitro* systems as compared to their *in vivo* analogs; how great are the changes caused by the introduction into a cul-

Correspondence: Dr. N. Leigh Anderson, Large Scale Biology Corporation, 9620 Medical Center Drive, Rockville, MD 20850, USA

Abbreviations: CBB, Coomassie Brilliant Blue; CPK, creatine phosphokinase; 2-D, two-dimensional; IEF, isoelectric focusing; MSN, master spot number; NP-40, Nonidet P-40; SDS, sodium dodecyl sulfate

0173-0835/91/1111-0907 $3.50+.25/0

ture and the associated shift to strong selection for growth, and how do these affect experimental outcomes? Hence the apparent advantages of *in vitro* systems, in terms of experimental manipulation, may be counterbalanced by other factors relating to 2-D data quality.

There is a second important class of reasons for exploring the use of an *in vivo* biological system such as the liver. Historically, there have been two broad approaches to the mechanistic dissection of biochemical processes in intact cellular systems: genetics (a search for informative mutants) and the use of chemical agents (drugs and chemical toxins). Both approaches help us to understand complex systems by disrupting some specific functional element and showing us the result. With the development of techniques for genetic manipulation and cloning, the genetic approach can be effectively applied either *in vitro* or *in vivo*, although the *in vitro* route is usually quicker. The chemical approach can also be applied to either sort of biological system; here, however, the bulk of consistently acquired information is in experimental animals (rats and mice). While most biologists know a short list of compounds having specific, experimentally useful effects (*e.g.*, inhibitors of protein synthesis, ionophores, polymerase inhibitors, channel blockers, nucleotide analogs, and compounds affecting polymerization of cytoskeletal proteins), there is a much larger number of interesting chemically-induced effects, most of them characterized by toxicologists and pharmacologists in rodent systems. Just as a thorough genetic analysis would involve saturating a genome with mutations, it is possible to imagine a saturating number of drugs, the analysis of whose actions would reveal the complete biochemistry of the cell. While organized drug discovery efforts usually target specific desired effects, the nature of the process, with its dependence on screening large numbers of compounds, necessarily produces many unanticipated effects. It is therefore reasonable to suppose that the required broad range of compounds necessary to achieve "biochemical saturation" may be forthcoming; in fact, it may already exist among the hundreds of thousands of compounds that failed to qualify as drugs.

Among organs, the liver is an obvious choice for the study of chemical effects because of its well-known plasticity and responsiveness. The brain appears to be quite plastic (*e.g.* [7]), but it is a complicated mixture of cell types requiring skillful dissection for most experiments. The kidney, while quite responsive, also presents a potentially confounding mixture of cell types. The liver, by contrast, is made up of one predominant cell type which is easy to solubilize: the hepatocyte, representing more than 95% of its mass. Most importantly, the liver performs many homeostatic functions that require rapid modulation of gene expression. It appears that most chemical agents tested affect gene expression in the liver at some dosage (N. Leigh Anderson, unpublished observations), an interesting contrast to our earlier work with lymphocytes, for example, which seem to be much less responsive. Such results conform to the expectation that cells with a homeostatic, physiological role should be more plastic than cells differentiated for a purpose dependent on the action of a limited number of specific genes.

The liver also allows the parallels between *in vitro* and *in vivo* systems to be examined in detail. Significant progress has been made in the development of mouse, rat and human hepatocyte culture systems, as well as in precision-cut tissue slices. Using such an array of techniques, it is possible to assemble a matrix of mammalian systems including mouse and rat *in vivo* on one level and mouse, rat and human *in vitro* on a second level, and to compare effects between species and between systems. This approach allows us to draw informed conclusions regarding the biochemical "universality" of biological responses among the mammals and to offer some insight into the validity of *in vitro* approaches for toxicological screening. We believe this data will be necessary if *in vitro* alternatives are to achieve wide usage in government-mandated safety testing of drugs, consumer products and industrial and agricultural chemicals.

A number of interesting studies have been published using 2-D mapping to examine effects in the rodent liver. A number of investigarors have made use of the technique to screen for existing genetic variants [8–11] or induced mutations [12–14], mainly in the mouse. This work builds on the wealth of genetic information available on the mouse and its established position as a mammalian mutation-detection system. While some studies of chemical effects have been undertaken in the mouse [15–17], most have used the rat [18–23]. The examination of the cytochrome p-450 system, in particular, has been carried out almost exclusively on the rat [24, 25].

These considerations lead us to conclude that rodent liver offers the best opportunity to systematically examine an array of gene regulation systems, and ultimately to build a predictive model of large-scale mammalian gene control. The basic underlying foundation of such a project is a reliable, reproducible master 2-D pattern of liver, to which ongoing experimental results can be referred. In this paper, we report such a master pattern for the acidic and neutral proteins of rat liver (pattern F344MST3). In future, this master will be supplemented by maps of basic proteins, and analogous maps of mouse and human liver.

2 Materials and methods

2.1 Sample preparation

Liver is an ideal sample material for most biochemical studies, including 2-D analysis. A sample is taken of approximately 0.5 g of tissue from the apical end of the left lobe of the liver. Solubilization is effected as rapidly as practical; a delay of 5–15 min appears to cause no major alteration in liver protein composition if the liver pieces are kept cold (*e.g.*, on ice) in the interim. In the solubilization process, the liver sample is weighed, placed in a glass homogenizer (*e.g.*, 15 mL Wheaton); 8 volumes of solubilizing solution*

* The solubilizing solution is composed of 2% NP-40 (Sigma), 9 M urea (analytical grade, *e.g.*, BDH or Bio-Rad), 0.5% dithiothreitol (DTT: Sigma) and 2% carrier ampholytes (pH 9–11 LKB: these come as a 20% stock solution, so 2% final concentration is achieved by making the final solution 10% 9–11 Ampholine by volume). A large batch of solubilizer (several hundred mL) is made and stored frozen at −80°C in aliquots sufficient to provide enough for one day's estimated sample preparation requirement. The solution is never allowed to become warmer than room temperature at any stage during preparation or thawing for use, since heating of concentrated urea solutions can produce contaminants that covalently modify proteins producing artifactual charge shifts. Once thawed, any unused solubilizer is discarded.

ded (*i.e.*, 4 mL per 0.5 g tissue) and the mixture is homogenized using first the loose- and then then the tight-fitting glass pestle. This takes approximately 5 strokes with each pestle and is carried out at room temperature because it would crystallize out in the cold. Once the liver sample is thoroughly homogenized in the solubilizer, it is assumed that all the proteins are denatured (by the chaotropic effect of the urea and NP-40 detergent) and the enzymes inactivated by the high pH (~9.5). Therefore these samples may be kept at room temperature until they can be centrifuged or frozen as a group (within several hours of preparation). The samples are centrifuged for 6×10^4 *g* min (*e.g.*, 500 000 *g* for 12 min using a Beckman TL-100 centrifuge). The centrifuge rotor is maintained at just below room temperature (*e.g.*, 15–20 °C), but not too cold, so as to prevent the precipitation of urea. The centrifuge of choice is a Beckman TL-100 because of the sample tube sizes available, but any ultracentrifuge accepting smallish tubes will suffice. When an appropriate centrifuge is not available near the site of sample preparation, samples can be frozen at −80 °C and thawed prior to centrifugation and collection of supernatants. Each supernatant is carefully removed following centrifugation and aliquoted into at least 4 clean tubes for storage. This is done by transferring all the supernatant to one clean tube, mixing this gently (to assure homogeneous composition) and then dividing it into 4 aliquots. The aliquots are frozen immediately at −80 °C. These multiple aliquots can provide insurance against a failed run or a freezer breakdown.

2.2 Two-dimensional electrophoresis

Sample proteins are resolved by 2-D electrophoresis using the 20 × 25 cm Iso-Dalt® 2-D gel system ([26–29]; produced by LSB and by Hoefer Scientific Instruments, San Francisco) operating with 20 gels per batch. All first-dimensional isoelectric focusing (IEF) gels are prepared using the same single standardized batch of carrier ampholytes BDH 4–8A in the present case, selected by LSB's batch-testing program for rat and mouse database work**). A 10 µL sample of solubilized liver protein is applied to each gel, and the gels are run for 33 000 to 34 500 volt-hours using a progressively increasing voltage protocol implemented by a programmable high-voltage power supply. An Angelique™ computer-controlled gradient-casting system (produced by LSB) is used to prepare second-dimensional sodium dodecyl sulfate (SDS) polyacrylamide gradient slab gels in which the top 5% of the gel is 11%T acrylamide, and the lower 95% of the gel varies linearly from 11% to 18%T.

This system has recently been modified so as to employ a commercially available 30.8%T acrylamide/*N,N'*-methylenebisacrylamide prepared solution (thus avoiding the handling of the solid acrylamide monomer) and three additional stock solutions: buffer (made from Sigma pre-set Tris), persulfate and *N,N,N',N'*-tetramethylethylenediamine (TEMED). Each gel is identified by a computer-printed filter paper label polymerized into the lower left corner of the gel. First-dimensional IEF tube gels are loaded directly (as extruded) onto the slab gels without equilibration, and held in place by polyester fabric wedges (Wedgies™, produced by LSB) to avoid the use of hot agarose. Second-dimensional slab gels are run overnight, in groups of 20, in cooled DALT tanks (10 °C) with buffer circulation. All run parameters, reagent source and lot information, and notations of deviation from expected results are entered by the technician responsible on a detailed, multi-page record of the experiment.

2.3 Staining

Following SDS-electrophoresis, slab gels are stained for protein using a colloidal Coomassie Blue G-250 procedure in covered plastic boxes, with 10 gels (totalling approximately 1 L of gel) per box. This procedure (based on the work of Neuhoff [30, 31]) involves fixation in 1.5 L of 50% ethanol and 2% phosphoric acid for 2 h, three 30 min washes, each in 2 L of cold tap water, and transfer to 1.5 L of 34% methanol, 17% ammonium sulfate and 2% phosphoric acid for 1 h, followed by the addition of a gram of powdered Coomassie Blue G-250 stain. Staining requires approximately 4 days to reach equilibrium intensity, whereupon gels are transferred to cool tap water and their surfaces rinsed to remove any particulate stain prior to scanning. Gels may be kept for several months in water with added sodium azide. The water washes remove ethanol that would dissolve the stain (and render the system noncolloidal, with high backgrounds). The concentrated ammonium sulfate and methanol solution is diluted by equilibration with the water volume of the gels to automatically achieve the correct final concentrations for colloidal staining. Practical advantages of this staining approach can be summarized as follows: (i) the low, flat background makes computer evaluation of small spots (max OD < 0.02) possible, especially when using laser densitometry; (ii) up to 1500 spots can be reliably detected on many gels (*e.g.*, rat liver) at loadings low enough to preserve excellent resolution; and (iii) reproducibility appears to be very good: at least several hundred spots have coefficients of reproducibility less than 15%. This value is at least as good as previous CBB methods, and significantly better than many silver stain systems.

2.4 Positional standardization

The carbamylated rabbit muscle creatine phosphokinase (CPK) standards [32] are purchased from Pharmacia and BDH. Amino acid compositions, and numbers of residues present in proteins used for internal standardization, are taken from the Protein Identification Resource (PIR) sequence database [33].

2.5 Computer analysis

Stained slab gels are digitized in red light at 134 micron resolution, using either a Molecular Dynamics laser scanner (with pixel sampling) or an Eikonix 78/99 CCD scanner. Raw digitized gel images are archived on high-density DAT tape (or equivalent storage media) and a greyscale videoprint prepared from the raw digital image as hard-copy backup of the gel image. Gels are processed using the Kepler® software system (produced by LSB), a commercially available workstation-based software package built on

** This material (succeeding certified batches of which are available from Hoefer Scientific Instruments) has the most linear pH gradient produced by any ampholyte tested except for the Pharmacia wide range (which has an unacceptable tendency to bind high-molecular weight acidic proteins, causing them to streak).

some of the principles of the earlier TYCHO system [34–41]. Procedure PROC008 is used to yield a spotlist giving position, shape and density information for each detected spot. This procedure makes use of digital filtering, mathematical morphology techniques and digital masking to remove the background, and uses full 2-D least-squares optimization to refine the parameters of a 2-D Gaussian shape for each spot. Processing parameters and file locations are stored in a relational database, while various log files detailing operation of the automatic analysis software are archived with the reduced data. The computed resolution and level of Gaussian convergence of each gel are inspected and archived for quality control purposes.

Experiment packages are constructed using the Kepler experiment definition database to assemble groups of 2-D patterns corresponding to the experimental groups (*e.g.*, treated and control animals). Each 2-D pattern is matched to the appropriate "master" 2-D pattern (pattern F344MST3 in the case of Fischer 344 rat liver), thereby providing linkage to the existing rodent protein 2-D databases. The software allows experiments containing hundreds of gels to be constructed and analyzed as a unit, with up to 100 gels displayed on the screen at one time for comparative purposes and multiple pages to accommodate experiments of > 1000 gels. For each treatment, proteins showing significant quantitative differences *vs.* appropriate controls are selected using group-wise statistical parameters (*e.g.*, Student's t-test, Kepler[®] procedure STUDENT). Proteins satisfying various quantitative criteria (such as $P < 0.001$ difference from appropriate controls) are represented as highlighted spots onscreen or on computer-plotted protein maps and stored as spot populations (*i.e.*, logical vectors) in a liver protein database. Quantitative data (spot parameters, statistical or other computed values) are stored as real-valued vectors in the database. Analysis of coregulation is performed using a Pierson product-moment correlation (Kepler procedure CORREL) to determine whether groups of proteins are coordinately regulated by any of the treatments. Such groups can be presented graphically on a protein map, and reported together with the statistical criteria used to assess the level of coregulation. Multivariate statistical analysis (*e.g.*, principal components' analysis) is performed on data exported to SAS (SAS Institute).

2.6 Graphical data output

Graphical results are prepared in GKS and translated within Kepler[®] into output for any of a variety of devices. Linedrawing output is typically prepared as Postscript and printed on an Apple Laserwriter. Detailed maps presented here have been generated using an ultra-high-resolution Postscript-compatible Linotronic output device. Greyscale graphics are reproduced from the workstation screen using a Seikosha videoprinter. Patterns are shown in the standard orientation, with high molecular mass at the top and acidic proteins to the left.

2.7 Experiment LSBC04

In the study described here 12-week-old Charles River male F344 rats were used. Diets were prepared at LSB, based on a Purina 5755M Basal Purified Diet. Lovastatin and cholestyramine were obtained as prescription pharmaceuticals, ground and mixed with the diet at concentrations of 0.075% and 1%, respectively. The high cholesterol diet was Purina 5801M-A (5% cholesterol plus 1% sodium cholate in the control diet). Animal work was carried out by Microbiological Associates (Bethesda, MD). Animals were acclimatized for one week on the control diet, fed test or control diets for one week, and sacrificed on day 8. Average daily doses of lovastatin and cholestyramine in appropriate groups were 37 mg/kg/day and 5 g/kg/day, respectively, based on the weight of the food consumed. Liver samples were collected and prepared for 2-D electrophoresis according to the standard liver protocol (homogenization in 8 volumes of 9 M urea, 2% NP-40, 0.5% dithiothreitol, 2% LKB pH 9–11 carrier ampholytes, followed by centrifugation for 30 min at 80 000 × *g*). Kidney, brain and plasma samples were frozen. Gels were run as described above, and the data was analyzed using the Kepler[®] system. Gels were scaled, to remove the effect of differences in protein loading, by setting the summed abundances of a large number of matched spots equal for each gel (linear scaling).

3 Results and discussion

3.1 The rat liver protein 2-D map

F344MST3 is a standard 2-D pattern of rat liver proteins, based on the Fischer 344 strain. This pattern was initiated from a single 2-D gel and extensively edited in an experiment comparing it to a range of protein loads, so as to include both small spots and well-resolved representations of high-abundance spots. More than 700 rat liver 2-D patterns have been matched to F344MST3 in a series of drug effects and protein characterization experiments, and numerous new spots (induced by specific drugs, for instance) have been added as a result. A modified version including additional spots present in the Sprague-Dawley outbred rat has also been developed (data not shown). Figure 1 shows a greyscale representation and Fig. 2 a schematic plot of the master pattern. More than 1200 spots are included, most of which are visible on typical gels loaded with 10 µL of solubilized liver protein prepared by the standard method and stained with colloidal Coomassie Blue. Master spot numbers (MSN's) have been assigned to all proteins, and appear in the following figures, each showing one quadrant of the pattern. Figure 3 shows the upper left (acidic, high molecular mass) quadrant, Fig. 4 the upper right (basic, high molecular mass) quadrant, Fig. 5 the lower left (acidic, low molecular mass) quadrant, and Fig. 6 the lower right (basic, low molecular mass) quadrant. The quadrants overlap as an aid to moving between them. The gel position (in 100 micron units), isoelectric point (relative to the CPK internal p*I* standards) and SDS molecular mass (from the calibration curve in Fig. 8) are listed for each spot (Table 1). Because of the precision of the CPK-p*I* values, these parameters can be used to relate spot locations between gel systems more reliably than using p*I* measurements expressed as pH. A major objective of current studies is the identification of all major spots corresponding to known liver proteins, as well as rigorous definitions of subcellular organelle contents. Of particular interest to us is the parallel development of identifications in the rat and mouse liver maps, allowing detailed comparisons of gene expression effects in the two systems. The results of these studies will be presented systematically in a later edition of this database.

we include here a useful series of 22 orienting identifications as an aid to other users of the rat liver pattern (Table

2 Carbamylated charge standards, computed p*I*'s and molecular mass standardization

We have previously shown that the use of a system of close-spaced internal p*I* markers (made by carbamylating a basic protein) offers an accurate and workable solution to the problem of assigning positions in the p*I* dimension [32]. The same system, based on 36 protein species made by carbamylating rabbit muscle CPK, has been used here to assign p*I*'s to most rat liver acidic and neutral proteins. The standards were coelectrophoresed with total liver proteins, and the standard spots added to a special version of the master pattern F344MST3. The gel *X*-coordinates of all liver protein spots lying within the CPK charge train were then transformed into CPK p*I* positions by interpolation between the positions of immediately adjacent standards (Table 1) using a Kepler[2] vector procedure.

It has proven possible to compute fairly accurate p*I* values for many proteins from the amino acid composition [42]. We have attempted here to test a further elaboration of this approach, in which we computed p*I*'s for the CPK standards themselves, based on our knowledge of the rabbit muscle CPK sequence and the fact that adjacent members of the charge train typically differ by blockage of one additional lysine residue (Table 3). We compared these values to similar computed p*I*'s for an additional set of carbamylated standards made from human hemoglobin beta chains and a series of rat liver and human plasma proteins of known position and sequence (Fig. 7, Table 4). The result demonstrates good concordance between these systems. Two proteins show significant deviations: liver fatty-acid binding protein (FABP; #1 in Table 4) and protein disulphide isomerase (#20 in the table). The FABP spot present on F344MST3 may represent a charge-modified version of a more basic parent spot closer to the expected p*I*, not resolved in the IEF/SDS gel. Of particular importance is the fact that, by comparing computed p*I*'s of sequenced but unlocated proteins with the CPK p*I*'s, we can assign a probable gel location without making any assumptions regarding the actual gel pH gradient. This offers a useful shortcut, given the vagaries of pH measurement on small diameter IEF gels. We have used this approach to compute the CPK p*I*'s of all rat and mouse proteins in the PIR sequence database, as an aid to protein identification (data not shown).

In order to standardize SDS molecular weight (SDS-MW), we have used a standard curve fitted to a series of identified proteins (Fig. 8). Rather than using molecular mass *per se*, we have elected to use the number of amino acids in the polypeptide chain, as perhaps a better indication of the length of the SDS-coated rod that is sieved by the second dimension slab. The resulting values were multiplied by 112 (the weighted average mass of amino acids in sequenced proteins) to give predicted molecular masses. Because we use gradient slabs, we have not constrained the fitting curve to conform to any predetermined model; rather we tried many equations and selected the best using the program "Tablecurve" on a PC. The equation chosen was $y = a + bx + c/x^2$, where y is the number of residues, x is the gel *Y* coordinate, a is 511.83, b is −0.2731 and c is 33183801. The resulting fit appears to be fairly good over a broad range of molecular mass.

3.3 An example of rat liver gene regulation: Cholesterol metabolism

Experiment LSBC04 was designed as a small-scale test of the regulation of cholesterol metabolism *in vivo* by three agents included in the diet: lovastatin (Mevacor®, an inhibitor of HMG-CoA reductase); cholestyramine (a bile acid sequestrant that has the effect of removing cholesterol from the gut-liver recirculation); and cholesterol itself. The first two agents should lower available cholesterol and the third should raise it, allowing manipulation of relevant gene expression control systems in both directions. Such an experiment offers an interesting test of the 2-D mapping system since most of the pathway enzymes are present in low abundance, many are membrane-bound and difficult to solubilize, and the pathway itself is complex. Approximately 1000 proteins were separated and detected in liver homogenates. Twenty-one proteins were found to be affected by at least one treatment, and these could be divided into several coregulated groups.

3.3.1 MSN 413 (putative cytosolic HMG-CoA synthase) and sets of spots regulated coordinately or inversely

One group of spots (including a spot assigned to the cytosolic HMG-CoA synthase, MSN 413) showed the expected increase in abundance with lovastatin or cholestyramine, the synergistic further increase with lovastatin and cholestyramine, and a dramatic decrease with the high cholesterol diet. Spot number 413 is the most strongly regulated protein in the present experiment, showing a 5- to 10-fold induction after a 1 week treatment with 0.075% lovastatin and 1% cholestyramine in the diet (Figs. 9 and 10). Its expression follows precisely the expectation for an enzyme whose abundance is controlled by the cholesterol level; it is progressively increased from the control levels by cholestyramine, lovastatin and lovastatin plus cholestyramine, and it sinks below the threshold of detection in animals fed the high cholesterol diet. This spot has been tentatively identified as the cytosolic HMG-CoA synthase, based on a reaction with an antiserum to that protein provided by Dr. Michael Greenspan at Merck Sharp & Dohme Research Laboratories. This enzyme lies immediately before HMG-CoA reductase in the liver cholesterol biosynthesis pathway, and is known to be co-regulated with it. Spot 413 has an SDS molecular weight of about 54 000 and a CPK p*I* of −11.4, in reasonably close agreement with a molecular weight of 57300 and a CPK p*I* of −15.7 computed from the known sequence of the hamster enzyme [43].

Using a classical product-moment correlation test (Kepler procedure CORREL), a series of five additional spots was found to be coregulated with 413. The level of correlation was exceedingly high (> 95%). Two of these, 1250 and 933, are at similar molecular weights and approximately one charge more acidic than 413 (Fig. 9), indicating that they may be covalently modified forms of the 413 polypeptide. This suspicion is strengthened by the observation that both spots are also stained by the antibody to cytosolic HMG-CoA synthase. The remaining three correlated spots appear